

# DIALOG와 인터넷 데이터베이스의 검색 효율성에 관한 비교 연구\*

## Internet Database Retrieval Efficiency vs. DIALOG Retrieval Efficiency

김현희(Hyun-Hee Kim)\*\*, 최창석(Chang-Seok Choi)\*\*\*,  
안태경(Tae-Kyoung Ahn)\*\*\*\*, 신명조(Myong-Cho Shin)\*\*\*\*\*

### 목 차

1 서 론	3.2 검색 효율성 비교
2 관련연구	3.2.1 가설
3 DIALOG와 인터넷 데이터베이스 비교	3.2.2 검색 성능 실험
3.1 기초조사	3.2.3 실험결과와 데이터 분석
3.1.1 데이터베이스 이용 행태 조사	4 결 론
3.1.2 특성비교	

### 초 록

본 연구에서는 정보시스템과 정보의 유형이 DIALOG와 인터넷 데이터베이스의 검색 효율에 어떤 영향을 미치는 지를 알아 보기 위해서 두 가지 작업을 수행하였다. 첫째는 이 두 시스템의 검색 성능에 대한 기초 자료를 얻기 위해서 두 시스템의 특성을 비교/분석하고 설문지를 통해 에너지 및 경제 분야 정보 이용자들의 정보 이용 행태를 조사하였다. 두번째는 에너지와 경제 분야의 20개의 정보 질문들을 이용하여 탐색 실험을 하여 두 시스템의 검색 효율성을 측정/비교하였는데 검색 효율성을 좀 더 효과적으로 비교하기 위해서 9가지 가설들을 설정하였다.

연구 결과를 종합해 보면 정보 유형이 DIALOG와 인터넷 데이터베이스 시스템의 적합성에 영향을 미칠 것이라는 가설들은 별 무리없이 검증되었으나 DIALOG를 활용한 검색 결과의 적합도가 인터넷을 활용한 경우의 적합도 보다 높을 것이라는 가설은 경제 분야만 통계적으로 검증되고 에너지 분야는 검증되지 못했다. 본 연구 결과에서 상대적으로 인터넷의 검색 효율이 높게 나온 이유로는 먼저 인터넷 검색에서 專門 정보가 비교적 많이 수록된 專門 검색엔진을 사용하게 한 점을 들 수 있고, 둘째는 인터넷 데이터베이스의 질적인 성장을 들 수 있으며, 끝으로 경제 및 에너지 분야의 주제 특성상 인터넷 데이터베이스를 통해 관련 자료를 많이 얻을 수 있지 않았을까 하는 가정을 해 볼 수 있다.

### ABSTRACT

This study compared finding economic and energy information on the WWW to finding the same information on DIALOG, a traditional search service. Professional searchers answered 20 questions for end users using either of DIALOG and one Internet database (general search engine or Web database). The relevance of the results in both sets of answers was ranked by searchers and end-users, respectively.

The study found that searching for information on the Web took at least twice as long as it did when using DIALOG. Relevance rating was a little higher for materials found on DIALOG. However, the relevance rating difference between two systems was not so higher than we expected. From the research results, we conclude that Internet database including Web database and general search engines is providing valuable information of economic and energy subject areas.

키워드: DIALOG, 인터넷, 적합성, 경제, 에너지, 인터넷 專門 검색엔진, 검색 효율성,  
인터넷 데이터베이스, 정보 유형, 인터넷 일반 검색엔진

\* 이 연구는 일부 정보통신부의 정보통신 우수시범학교 지원사업에 의하여 수행된 것임

\*\* 명지대학교 문헌정보학과/정보통신교육연구센터

\*\*\* 명지대학교 전자정보통신공학부/정보통신교육연구센터

\*\*\*\* 대외경제정책연구원

\*\*\*\*\* 한국가스공사

■ 논문 접수일 : 2000년 2월 20일

## 1 서 론

### 1.1 연구 목적

인터넷 데이터베이스를 통해 검색되고 있는 자료들이 수 년 전만 해도 비전문 자료가 주류를 이루었으나 최근에 와서는 상용 온라인 벤더와 거의 같은 질의 정보를 탐색할 수 있거나 심지어는 온라인 벤더에서 탐색할 수 없는 정보를 제공하기도 한다.

인터넷 데이터베이스는 많은 정보검색사들에게 유용한 자료를 제공함으로써 상용 온라인 벤더의 대안으로서 그 중요성이 커지고 있는 반면 아직까지 일부 정보검색사들은 인터넷 데이터베이스가 탐색 시간이 길고 정보의 질에 일관성이 없다는 주장을 하기도 한다. 이와 같이 정보검색사간의 의견이 다른 것은 정보 유형에 따라 온라인 벤더와 인터넷 데이터베이스간의 정보 질에 많은 차이가 생길 수 있기 때문이라고 가정해 볼 수 있다.

본 연구에서는 정보시스템과 정보의 유형이 DIALOG와 인터넷 데이터베이스의 검색 효율에 어떤 영향이 미치는 지를 알아 보기 위해서 두 가지 작업을 수행한다. 첫째는 이 두 시스템의 검색 성능에 대한 기초 자료를 얻기 위해서 두 시스템의 특성을 비교/분석하고 설문지를 통해 에너지 및 경제 분야 정보 이용자들의 정보 이용 행태를 조사한다. 두번째는 에너지와 경제 분야의 정보 질문들을 이용하여 탐색 실험을 한 후 두 시스템의 검색 효율성을 측정/비교한다. 본 연구 결과는 에너지 및 경제 분야에 관여하는 정보 검색사/사서와 최종 이용자들이 좀 더 효율적인 정보 검색에 이용될 수 있을 뿐만 아니라 데이터베이스 선정 도구의 지식베이스의 자료로 활용될 수 있다.

### 1.2 연구 방법과 범위

본 연구는 주제분야를 에너지와 경제로 제한하여 이 두 분야의 인터넷 데이터베이스와 DIALOG의 검색 효율성을 비교 분석하기 위해서 다음과 같은 절차를 이용하여 자료를 수집하고 검색 실험을 수행한다.

첫째, DIALOG와 인터넷 데이터베이스의 검색 성능에 대한 기초 자료를 얻기 위해서 설문지 조사와 두 시스템간의 특성을 분석한다. 먼저 설문지 조사는 경제 및 에너지 분야의 이용자들의 데이터베이스 이용 행태를 파악하기 위해서 연구원, 사서 등 에너지 및 경제경영 분야에 종사하고 있는 100명의 이용자들에게 설문지를 전자우편과 일반우편을 이용하여 배포하여 에너지분야 27명, 경제분야 33명, 총 70명으로부터 설문지를 회수하였다. 다음으로 시스템 매뉴얼, 일반 문헌, 데이터베이스 탐색 및 실제 테스트를 통하여 인터넷 데이터베이스와 DIALOG의 특성을 고찰한다.

둘째, 두 시스템간의 검색 효율을 적합성, 탐색 소요시간, 비용의 측면에서 비교하기 위해서 검색 실험을 수행한다. 적합성 측정에 앞서 정보시스템 유형과 정보 유형을 독립 변인으로, 적합도를 종속변인으로 세운 총 9개의 가설을 설정한다. 검색 실험은 먼저 5개의 에너지 및 경제 관련 연구기관에서 연구자들이 일반적으로 요청하는 질문 중 20개(에너지분야 10개, 경제 분야 10개)를 추출한다. 그 다음 3년 이상의 경력을 갖고 있는 2명의 정보검색사에게 의뢰하여 두 검색시스템을 이용하여 탐색을 하게 한 후 먼저 탐색소요시간(정보 검색과 가공 시간)과 비용을 비교 분석한 후 끝으로 2명의 정보 이용자(검색사와 질문자)에게 최종 적합성을 평가받아 가설 검증을 한다.

### 1.3 용어 정의

본 연구에서 사용된 용어의 개념을 정의하면 다음과 같다.

(1) 인터넷 일반 검색엔진: 전통적인 인터넷 체 공간, 즉 HTTP, Gopher, Usenet, FTP 등에 저장되어 있는 내부 문서, 파일들을 색인하여 구축한 데이터베이스이다. Yahoo, AltaVista 등이 여기에 해당된다.

(2) 인터넷 專門 검색엔진: 인터넷에 연결되어 있는 외부문서, 파일들을 색인하여 구축한 주제 별로 구성된 데이터베이스이다. EDGAR, QPAT, Techweb 등이 여기에 해당된다.

(3) 인터넷 데이터베이스: 인터넷 일반 검색엔진과 專門 검색엔진을 포괄적으로 지칭하는 말이다.

## 2 관련 연구

### 2.1 특징 비교

라찬스(LaChance, 1998)는 인터넷과 전통적인 온라인 시스템을 비교 분석한 후 이 두 시스템의 장단점들을 정리하였다. 베이츠(Bates, 1999)는 비즈니스/경제 분야의 11명의 정보검색사와 면담한 내용을 정리하였는데 검색자들이 느끼는 인터넷과 온라인 벤더의 검색 기능의 차이에 대해 설명하고 있다.

橋田(1998)은 검색엔진에 의한 인터넷 탐색과 상용데이터베이스 탐색을 수록정보, 색인파일, 명령어체계, 결과의 표시방법의 네가지 관점에서 비교하였다. 추(Chu, 1998)는 온라인 검색 서비스와 인터넷 검색 서비스의 차이점을 데이터베이스 생성과 구조, 탐색 기능 및 이용자의 노력의

세가지 측면에서 서술하였다.

이러한 특징 비교 연구가 있으나 대부분의 연구가 인터넷 일반 검색엔진에 국한하여 온라인 벤더와의 특성을 비교한 것이며 많은 專門 정보를 수록하고 있는 인터넷 專門 검색엔진과의 특성을 비교한 연구는 거의 없었다. 따라서 인터넷 일반 검색엔진과 인터넷 專門 검색엔진을 포함한 인터넷 데이터베이스와 온라인 벤더와의 특성 비교가 필요하다고 사료된다.

### 2.2 검색 효율 비교

펠드만(Feldman, 1998)은 인터넷 일반 검색엔진과 전통적인 상용온라인 검색시스템을 비교하였다. 이 연구에서 정보검색사는 두 개의 온라인 서비스(DIALOG와 Dow Jones Interactive)와 적어도 하나 이상의 웹 검색엔진을 사용하여 정보 검색을 실시한 후 검색 결과를 분석하여 이 두 시스템의 검색 효율성을 평가하였다. 김현희(1999)는 인터넷 일반 검색엔진과 인터넷 專門 검색엔진을 사용한 인터넷 데이터베이스와 온라인 시스템(PQD, EBSCOhost, First Search)과의 검색 효율성을 비교 분석하였다.

첫번째, 펠드만의 연구는 경제 및 에너지 분야를 특별히 다루지 않았을 뿐 만 아니라 인터넷 일반 검색엔진에 국한하여 수행한 연구이다. 또한 이 연구는 외국에서 행해진 연구이기 때문에 연구 결과를 정보 및 네트워크 환경이 다른 국내에 그대로 적용하는 데는 어려움이 따른다. 두 번째 김의 연구는 주제를 비즈니스로 국한했을 뿐 아니라 실험 온라인 데이터베이스로 전통적인 온라인 벤더 대신 인터넷 專門 검색엔진에 가까운 EBSCOhost 등을 활용하고 있다.

이에 따라 국내 정보 환경하에서 전통적인 온

라인 벤더와 인터넷 일반 검색엔진과 인터넷 專門 검색엔진을 포함한 인터넷 데이터베이스간의 좀 더 자세한 비교 연구가 필요한 시점이라 사료된다.

### 3 DIALOG와 인터넷 데이터베이스 비교

#### 3.1 기초 조사

##### 3.1.1 데이터베이스 이용 행태조사

두 시스템을 비교하는 작업에 앞서 DIALOG와 인터넷 데이터베이스간의 검색 효율에 대한 기초 자료를 얻기 위해서 경제 및 에너지 분야의 이용자들의 상용 데이터베이스와 인터넷의 이용 현황을 파악하기 위해서 설문지 조사를 먼저 수행하였다. 연구원, 사서 등 기술 및 경제경영 분야에 종사하고 있는 100명의 이용자들에게 설문지를 전자우편과 일반우편을 이용하여 배포하여 70명(에너지분야: 27명, 경제분야: 33명)으로부터 설문지를 회수하였다. 응답자 대부분이 석사 학위 이상의 소지자였다. 다음은 설문지 분석 내용 중 중요한 자료들을 정리한 내용이다.

에너지와 경제 분야 연구자들이 가장 많이 사용하는 국내 및 해외 검색엔진은 야후코리아(에너지 27%, 경제 36%)와 알타비스타(에너지 17%, 경제 20%)로 나타났다. 에너지 분야에서 인터넷에서 자주 이용하는 정보는 기관/회사정보(21%), 학술논문/서지데이터(18%), 생산품 정보(12%) 순인데 반해 경제 분야에서는 최신정보(31%), 통계/수치(23%), 기관 및 회사정보(11%) 순이었다. 에너지 분야에서 인터넷 검색시 가장 불만스러운 점은 專門정보의 부족(25%)을 들었으며 경제 분야에서는 최신정보만 수록한 점(22%)을 지적하였다.

에너지분야에서 해외 온라인 벤더에서 자주 이용

하는 정보는 학술논문/서지데이터(22%), 기관/회사정보(20%), 학술회의 행사정보(17%) 순으로 나타난 반면 경제 분야에서는 학술 논문의 쏠문(22%), 특허정보(16%), 학술논문/서지데이터(14%) 순이었다. 에너지 분야에서 사용하는 온라인 벤더는 DIALOG(31%), JOIS(22%)순으로 나타났다. 경제학 분야는 ProQuest Direct(30%), DowJones Interactive(19%)순으로 나타났다.

에너지 분야에서는 온라인 벤더 이용시 71%의 이용자들이 사서/검색사에게 의뢰하고 7%의 이용자들만 직접 탐색한 반면 경제 분야에서는 29%의 이용자만 전문가에게 의뢰하고 47%의 연구자들이 직접 탐색을 수행하였다. 이는 에너지 분야에서 사용하는 온라인 벤더들의 사용법이 경제 분야에서 사용하는 온라인 벤더들의 사용법보다 어려운 것이 한 원인으로 사료된다.

#### 3.1.2 특성 비교

##### 3.1.2.1 분석 도구

DIALOG와 인터넷 시스템의 특징을 비교하기 위해 먼저 관련 문헌을 분석하고 아울러 시스템의 매뉴얼, 천리안의 도원컨설팅에서 제공하는 자료 그리고 각 시스템을 실제 이용해 보았다.

##### 3.1.2.2 비교 결과

인터넷에서 일반 검색엔진과 專門 검색엔진의 특성은 상당한 차이가 나기 때문에 이를 세분하여 <표 1>에 보여주는 것 처럼 온라인 벤더인 DIALOG와 인터넷 일반 검색엔진 및 인터넷 專門 검색엔진의 세 시스템의 특성을 비교하였다.

세 시스템의 중요한 특성 즉 검색 공간, 데이터베이스 생성과 구조, 탐색 기능, 탐색 결과의 등급 표시 및 중복 자료 처리에 대해 자세히 설명하면 다음과 같다.

〈표 1〉 DIALOG와 인터넷 데이터베이스의 특성 비교

구 분	DIALOG	인터넷 데이터베이스	
		일반 검색엔진	專門 검색엔진
데이터베이스별 보유자료수	2백만 미만의 레코드를 보유함	5십만개-1억 4천만개 웹문서	만개-5천만개
데이터베이스 및 서비스수	5백개 이상	2천개 이상	십만개 이상
검색공간	개별 및 다중 데이터베이스 탐색가능	인터넷 내부 문서 검색	인터넷 외부 문서 연동
색인 범위	서지정보(표제, 저자 등), 디스크립터, 초록 등	쵸文, 서명, 표목, URL, 요약/초록, 메타태그, 낱짜 등	메타 데이터
데이터베이스 생성과 구조	문헌수집과 디스크립터 할당은 수작업으로 하나 색인 작업은 자동으로 함. 데이터베이스는 정형화된 레코드로 구성됨	문헌수집과 색인 작업을 자동(또는 수동)으로 함. 데이터베이스는 비정형화된 문서들로 구성됨	대부분 데이터베이스는 정형화된 레코드로 구성됨. 색인은 수동으로 함
탐색 기능	블리언연산 탐색, 절단검색, 인접연산 검색, 필드검색, 구 검색, 통제어/자연어 검색, 가중치 검색 등 가능	블리언연산 검색, 절단 검색, 인접연산 검색, 제한된 필드 검색, 구 검색, 자연어 탐색, 가중치 검색 등 가능	블리언연산 검색, 절단검색, 인접연산 탐색, 필드검색, 구 검색, 유의어 검색 등 기능
검색결과외 등급표시	유	유	무/유
갱신율	계획된 갱신 스케줄을 갖고 있음	계속적으로 갱신되지만 불규칙한 경우가 많음	계속적으로 갱신되지만 불규칙한 경우가 많음
자료 일관성	자료의 질에서 일관성이 있음	자료의 질에서 일관성이 없음	자료의 질에서 일관성이 있음
타겟 대상 및 분야	훈련된 정보 검색사를 대상으로 하며 비즈니스, 교육 및 연구 분야에 초점을 맞춤	최종 이용자 대상이며 생활 정보와 같이 최종 이용자들이 관심있는 분야들에 초점을 맞춤	최종 이용자 대상이며 학술 정보 및 비즈니스 분야 등에 초점을 맞춤
언 어	영어	영어를 주로 지원하나 일어, 한국어, 중국어를 지원하는 엔진도 있음	영어 중심
요 금	유료	무료	무료(90%), 유료(10%)
중복 자료 처리	자동 처리	많은 노력을 요함	중복 자료 걸러냄

### 1) 검색 공간

DIALOG의 경우 한번에 하나의 데이터베이스를 탐색할 수 있을 뿐 아니라 40개 이상의 데이터베이스를 동시에 다중 탐색할 수 있다. 인터넷 일반 검색엔진의 검색 공간은 전통적인 인터넷 제 공간, 즉 HTTP, Gopher, Usenet, FTP 등에 저장되어 있는 내부 문서, 파일들을 색인하여 구축한 데이터베이스인 반면 인터넷 專門 검색엔진은 인터넷에 연결되어 있는 외부문서, 파일들을 색인하여 구축한 데이터베이스이다.

### 2) 데이터베이스 생성과 구조

DIALOG의 경우 문헌수집과 통제어 색인어인 디스크립터 추출은 수작업으로 하나 기본서지파일에서 색인어를 추출하여 색인파일을 구성할 때는 자동으로 한다. 이러한 기본서지 및 색인 파일은 레코드와 필드로 구성되며 계층적 구조를 갖는다. 인터넷 일반 검색엔진의 경우 AltaVista는 스파이더(spiders)에 의해 문헌이 자동 수집되며 색인 작업은 색인 프로그램에 의해 웹페이지 창조자에 의해 직접 제출된 사이트들과 스파이더에 발견된 사이트들이 자동으로 색인되는 반면 야후는 수작업으로 색인 작업을 한다. 인터넷 정보원은 HTML 이라는 마크업 언어를 사용하지만 문서 형식은 정해진 형식이 없이 비정형화되어 있으며 하이퍼텍스트나 하이퍼미디어 등의 네트워크 구조로 이루어진다. 인터넷 專門 검색엔진은 인터넷 일반 검색 엔진과는 달리 대부분의 정형화된 레코드로 구성되며 색인은 수작업 방식을 주로 채택하고 있다.

### 3) 탐색 기능

기본탐색능력에 있어 DIALOG와 인터넷 데이터베이스 모두 다 블리언 탐색과 절단 탐색 등의 기능을 제공하고 있으나 DIALOG가 좀 더 폭넓은 탐색방법을 제공하고 있다. 필드탐색의 경우

DIALOG는 비교적 정교한 필드탐색(제한탐색)을 허용하고 있는 반면 인터넷 일반 검색엔진은 극히 제한된 유사필드탐색(예, 표제 및 URL)만을 허용하고 있다.

통제어 탐색의 경우 DIALOG에서는 오래전부터 제공해온 기능이지만 인터넷 탐색에서는 몇몇 탐색 서비스에서 부분적이고 특수화된 분야를 제외하고는 대부분 자연어 탐색만 수행하고 있다. 가중치 탐색은 핵심 디스크립터에 의해서 DIALOG상에서 이루어지고 있고, 인터넷 일반 검색엔진(AltaVista)에서는 탐색어를 중요도 순서로 입력하게 함으로써 탐색어에 가중치를 부여할 수 있도록 한다.

### 4) 탐색 결과의 등급 표시

DIALOG의 경우 명령어 "rank"를 사용하여 검색 결과의 특정 필드를 지정하여 해당 필드안에서 개념이나 이름 등의 출현빈도에 따라 순위를 매기고 있다. 인터넷 검색엔진(AltaVista)인 경우 세가지 기준 즉 탐색어가 문헌에 출현한 빈도, 탐색어들의 인접도 및 탐색어의 출현 장소에 의해 탐색 결과를 등급지운다. 인터넷 專門 검색엔진도 대부분 탐색 결과를 등급화 해서 출력하고 있다.

### 5) 중복 자료 처리

탐색 결과를 편집하는 하나의 과정으로 중복된 자료를 추려내는 작업이 있다. 중복 방지 기능은 온라인상에서는 개개 및 다중 데이터베이스 탐색에서 지원을 하고 있다. 인터넷 일반 검색엔진의 탐색결과에서 생기는 문헌 중복은 인터넷 서비스 상에서 심각한 문제를 불러일으키고 있다. 즉 인터넷 이용자들은 탐색 결과에서 비적합 문헌외에 중복된 자료를 걸러내기 위해서 많은 시간을 낭비하고 있다. 인터넷 專門 검색엔진은 수작업으로 색인 작업을 하기 때문에 중복 자료는 정보 가공시 걸러진다.

### 3.2 검색 효율성 비교

#### 3.2.1 가 설

검색 실험에 앞서 관련 문헌을 분석하고 아울러 앞에서 작업한 두 시스템간의 특성 비교와 데이터베이스 이용 행태의 분석 결과를 기초로 하여 정보 시스템과 정보의 유형을 각각 독립 변인으로, 적합성을 종속변인으로 하여 가설을 설정하면 다음과 같다.

##### 3.2.1.1 가 설 1

정보시스템 유형(DIALOG와 인터넷 데이터베이스 시스템)이 적합성에 영향을 미칠 것이다.

- 1) 1-1. 에너지 분야에서 DIALOG를 활용한 검색 결과의 適合度가 인터넷을 활용한 경우의 적합도 보다 높을 것이다.
- 2) 1-2. 경제 분야에서 DIALOG를 활용한 검색 결과의 적합도가 인터넷을 활용한 경우의 적합도 보다 높을 것이다.

##### 3.2.1.2 가 설 2

정보 유형이 DIALOG와 인터넷 데이터베이스 시스템의 적합성에 영향을 미칠 것이다.

- 1) 2-1. 학술잡지의 全文 정보는 인터넷 데이터베이스 보다 DIALOG의 적합도가 더 높을 것이다.
- 2) 2-2. 학술잡지의 書誌 정보는 인터넷 데이터베이스 보다 DIALOG의 적합도가 더 높을 것이다.
- 3) 2-3. 특허 정보는 인터넷 데이터베이스 보다 DIALOG의 적합도가 더 높을 것이다.
- 4) 2-4. 수치 및 통계 정보는 DIALOG 보다 인터넷 데이터베이스의 적합도가 더 높을 것이다.
- 5) 2-5. 동향분석자료는 DIALOG 보다 인터넷 데이터베이스의 적합도가 더 높을 것이다.

6) 2-6. 회색문헌 및 회의 정보는 DIALOG 보다 인터넷 데이터베이스의 적합도가 더 높을 것이다.

7) 2-7. 비영어권 자료는 DIALOG 보다 인터넷 데이터베이스의 적합도가 더 높을 것이다.

#### 3.2.1 검색 성능 실험

##### 3.2.1.1 실험도구

본 연구의 실험대상도구인 인터넷 데이터베이스는 검색자가 자신이 주로 사용하는 탐색 도구를 사용하도록 하여 특별한 제한을 두지 않았다. 이는 검색자들이 주로 사용하고 있는 탐색도구가 어떠한 것인지, 주제분야별 질문에 어떠한 효율을 나타내는지를 조사하기 위해서이다. 한편 온라인 데이터베이스는 DIALOG로 한정하였는데 이는 설문지 분석 결과 DIALOG가 에너지 분야 이용자들이 가장 선호하면서 경제 분야의 상당수 이용자들도 이용하고 있는 벤더이며 또한 다양한 정보원을 확보하고 있는 세계 최대의 데이터베이스이기 때문이다.

##### 3.2.1.2 실험방법과 절차

실험방법은 에너지와 경제분야의 실험 질문을 각각 두 명의 3년 이상된 정보검색사에게 의뢰하였다. 질문은 각 분야에서 이용자들이 요구한 것으로 분야별로 각각 10개, 모두 20개의 질문을 인터넷 데이터베이스와 DIALOG에서 검색하도록 하였다.

성능평가에 사용된 기준은 검색시간과 가공시간을 포함한 탐색소요시간, 적합성, 비용 등을 사용하였다. 검색시간은 검색자가 정보를 검색하는 시간이며 가공시간<sup>1)</sup>은 검색된 자료를 이용자가 보기까지

1) 정보가공시간으로 검색된 정보를 중복자료, 페이지를 찾을 수 없는 자료, 적합성이 전혀 없는 자료 등을 제외하고 이용자가 볼 수 있게 파일을 저장하여 바로 볼 수 있게끔 가공하는데 걸리는 시간

〈표 2〉 질문1의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	IMF	AltaVista			Econbase ABI	ABI		
탐색도구 (DB명)	IMF	AltaVista			Econbase ABI	ABI		
최종검색건수	1	1	1	1	20	15	17.5	17.5
데이터베이스 이용료(\$)	-	-			74,434 (65.86)	45,467 (40.23)	59,950	59,950
검색시간(분)	5	10	7.5	5.5	4	3	3.5	2.6
가공시간(분)	3	4	3.5		2	1.5	1.7	
총비용	1,384	2,422	1,903	1,903	75,472	46,245	60,858	60,858
적합성1	5	3	4	3.2	2	3	2.5	2.2
적합성2	3	2	2.5		3	1	2	

의 소요된 시간이다. 적합성<sup>2)</sup>은 두 가지로 적합성1은 검색자가 판정한 적합성이고 적합성2는 질문자가 최종 판정한 적합성이다. 적합성 판정은 최종 검색 건수에서 20개 문헌을 추출하여 수행하였다. 최종 적합도는 검색자와 질문자가 판정한 적합도를 합하여 2로 나눈 평균값을 이용하였다. 비용<sup>3)</sup>은 인터넷의 경우 탐색소요시간을 기초 자료로 하여 일반적인 검색사 시간 당 비용을 산출하여 측정하였으며 DIALOG의 경우는 인터넷에서처럼 탐색소요시간을 기초 자료로 하여 검색사 시간 당 비용을 산출하여 측정된 비용에 다 데이터베이스 이용료를 포함시켰다.

인터넷의 경우 S1은 IMF 웹사이트에 직접 접속하였고, S2는 AltaVista를 경유하여 IMF 웹사이트에 접속하였다. DIALOG는 인터넷 보다 적합도가 낮았는데 이는 초록 또는 일부 기사만을 조사할 수 밖에 없었을 뿐 아니라 최신성도 유지하지 못하고 있었기 때문이었다.

② 질문 2: 일본 아시아 경제 연구소가 최근 발행한 2000년 동아시아 전망 관련 자료를 조사하고자 한다(비영어권).

인터넷의 경우 S1은 아시아경제연구소 웹사이트에 직접 접속하였고, S2는 AltaVista를 경유하여

3.2.2 실험결과와 데이터 분석

3.2.2.1 검색 결과

1) 경제 분야

① 질문 1: 현재(2000년)부터 2010년까지 한국(Korea)의 GDP 전망치(Economic Outlook)를 연간 시계열로 조사하고자 한다(통계/수치).

2) 검색결과에 대한 적합성으로 다음과 같은 구간 척도를 사용하였다.

- 0: (미해결)
- 1: 1-20(매우 부적합)
- 2: 21-40(비적합)
- 3: 41-60 (보통)
- 4: 61-80(비교적 적합)
- 5: 81-100(적합)

3) 정부고시 물가정보지 유통물가 '99년 9월 기준으로 정보검색사와 유사직종인 정보통신 기사 임금으로 측정



〈표 3〉 질문2의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	S1	S2			S1	S2		
탐색도구	아시아 경제 연구소	AltaVista			EconLit	Econbase		
최종검색건수	1	1	1	1	15	15	15	15
데이터베이스 이용료(\$)	-	-			42,230 (38.25)	34,019 (30.10)	38,624	38,624
검색시간(분)	2	10	6	4.5	3	2	2.5	1.8
가공시간(분)	3	3	3		1	1.5	1.2	
총비용	865	2,249	1,557	1,557	43,922	34,624	39,273	39,273
적합성1	5	5	5	5	0	0	0	0
적합성2	5	5	5		0	0	0	

아시아경제연구소 웹사이트에 접속하였다. DIALOG 검색에서는 EconLit, Econbase 모두 낮은 적합도를 나타내고 있는데 이는 비영어권 자료인 점과 최신성을 유지하고 있지 못했기 때문이다.

③ 질문 3: 동아시아의 경제발전에 관한 폴크 루그만과 제프리삭스의 논쟁에 대한 논문을 찾고자한다(書誌정보).

인터넷의 경우 S1은 아시아금융위기 웹사이트

〈표 4〉 질문3의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	S1	S2			S1	S2		
탐색도구	아시아 금융위기	AltaVista			EconLit	EconLit		
최종검색건수	15	15	15	15	15	15	15	15
데이터베이스 이용료(\$)	-	-			40,235 (35.60)	34,211 (30.27)	37,223	37,223
검색시간(분)	6	5	5.5	7.2	7	5	6	4
가공시간(분)	8	10	9		2	2	2	
총비용	2,422	2,595	2,508	2,508	41,792	35,422	38,607	38,607
적합성1	4.5	4	4.2	3.8	5	5	5	4.9
적합성2	4	3	3.5		4.8	4.9	4.8	

로 직접 접속하였고 S2는 AltaVista를 사용하여 검색한 결과 값이다. DIALOG 검색에서는 S1, S2 모두 EconLit을 사용하여 검색하였으며 높은

적합도를 나타내고 있다.

④ 질문 4: 아시아와 남미의 해외투자 비교에 관한 학술 논문을 조사하고자 한다(全文정보).

〈표 5〉 질문4의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	S1	S2			S1	S2		
탐색도구	AltaVista	AltaVista			EconLit	EconLit		
최종검색건수	15	15	15	15	7	14	10.5	10.5
데이터베이스 이용료(\$)	-	-			7,120 (6.30)	11,528 (10.2)	9,324	9,324
검색시간(분)	8	6	7	6.5	4	5	4.5	3.2
가공시간(분)	6	6	6		2	2	2	
총비용	2,422	2,076	2,249	2,249	8,158	12,739	10,448	10,448
적합성1	2.6	0	1.3	1.1	5	5	5	4.9
적합성2	2	0	1		5	4.9	4.9	

〈표 6〉 질문5의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	S1	S2			S1	S2		
탐색도구	IMD	AltaVista			ABI EconLit	EconLit		
최종검색건수	1	1	1	1	15	17	16	16
데이터베이스 이용료(\$)	-	-			71,349 (63.13)	34,064 (30.14)	52,706	52,706
검색시간(분)	4	5	4.5	3.7	6	4	5	3
가공시간(분)	3	3	3		1	1	1	
총비용	1,211	1,384	1,297	1,297	72,560	34,929	53,744	53,744
적합성1	5	5	5	5	3	1	2	1.5
적합성2	5	5	5		1	1	1	

인터넷의 경우 두 검색사 모두 AltaVista를 사용하였으며 각기 다른 키워드를 사용함으로써 양자간 적합도에 상당한 차이를 나타내고 있다. DIAOG 검색에서는 EconLit를 사용하였는데 모두 균등하게 높은 적합도를 나타내고 있다.

⑤ 질문 5: 1999년 한국의 국가경쟁력은 세계 몇 위인가(통계/수치).

인터넷 검색에서 S1 검색자는 IMD 웹사이트에 직접 접속하였고 S2는 AltaVista를 사용하였다. DIALOG 검색에서 S1은 ABI와 EconLit를 사용하여 검색하였고, S2는 EconLit만을 검색하였다. 적합도는 인터넷 검색의 경우 모두 全文으로 통계자료를 싣고 있어 높은 적합도를 나타내고 있으며 DIALOG 검색에서는 초록 또는 기사 검색으로 낮은 적합도를 나타내고 있다.

⑥ 질문 6: World Economic Forum에서 개최하는 2000년 연례회의의 중개 비디오 파일을 입수하고자 한다(회색문헌/회의정보).

인터넷 검색은 S1은 WEF(World Economic

Forum) 웹사이트에 직접 접속하여 비디오파일을 다운로드 받았다. S2는 AltaVista를 통해 WEF 사이트에 접속한 경우이다. DIALOG는 비디오 파일을 제공하지 않기 때문에 검색대상에서 제외하였다.

⑦ 질문 7: 우리나라의 해외국채 발행 및 수급에 관하여 조사하고자 한다(통계/수치).

인터넷 검색에서는 두 검색사 모두 AltaVista를 사용하였는데 적합한 통계자료를 검색하지 못했다. DIALOG 검색은 S1은 Bond Buyer, S2는 ABI, EconLit를 사용하였는데 Bond Buyer에서만 전문(full-text) 데이터를 입수할 수 있었다. 적합도는 양쪽 모두 낮은 것으로 나타났다.

⑧ 질문 8: 인도의 최근 경제 개황에 대하여 조사하고자 한다(동향분석정보).

인터넷 검색에서는 두 검색사 모두 AltaVista를 사용하였다. DIALOG 검색의 경우 S1은 EIU(Economist Intelligence Unit)를, S2는 EconLit를 검색하였는데 EIU의 경우 비교적 높

〈표 7〉 질문6의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	S1	S2						
탐색도구	WEF	AltaVista						
최종검색건수	1	15	8	8				
데이터베이스 이용료(\$)	-	-						
검색시간(분)	2	5	3.5	3.7				
가공시간(분)	2	6	4					
총비용	692	1,903	1,297	1,297				
적합성1	5	4.3	4.6	4.6				
적합성2	5	4.5	4.7					

은 적합도를 나타내었고 EconLit는 서지정보만 검색 할 수 있었다.

주가지수를 연도별 시계열로 조사하고자 한다 (통계/수치).

⑨ 질문 9: 다우존스와 니케이 지난 10년간

인터넷 검색에서는 두 검색사 모두 AltaVista

〈표 8〉 질문7의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
탐색도구	AltaVista	AltaVista			Bond Buyer	ABI EconLit		
최종검색건수	3	1	2	2	14	0	7	7
데이터베이스 이용료(\$)	-	-			52,531 (46.48)	1,966 (1.74)	27,248	27,248
검색시간(분)	10	10	10	7.2	2	1.1	1.5	1
가공시간(분)	6	3	4.5		1	0	0.5	
총비용	2,768	2,249	2,508	2,508	53,050	2,156	27,603	27,603
적합성1	2.7	1	1.8	1.4	2	0	1	0.7
적합성2	1.3	1	1.1		1	0	0.5	

〈표 9〉 질문8의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
탐색도구	AltaVista	AltaVista			EIU	EconLit		
최종검색건수	66	1	33.5	33.5	83	15	49	49
데이터베이스 이용료(\$)	-	-			454,351 (402.01)	22,886 (20.25)	238,618	238,618
검색시간(분)	3	10	6.5	5.2	2	4	3	2
가공시간(분)	5	3	4		1	1	1	
총비용	1,384	2,249	1,816	1,816	454,870	23,751	239,310	239,310
적합성1	2.8	3	2.9	2.7	5	2	3.5	3
적합성2	3	2	2.5		4	1	2.5	

를 사용하였지만 적절한 정보를 검색하지 못했다. DIALOG 검색에서 S1은 Bridge World Market News, S2는 ABI를 검색하였다. 적합도

는 일부 단편적인 기사정보 제공으로 매우 낮게 나타났다.

⑩ 질문 10: 미국의 1998년 지역별 해외투자

〈표 10〉 질문9의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	S1	S2			S1	S2		
탐색도구	AltaVista	AltaVista			Bridge World Market News	ABI		
최종검색건수	3	2	2.5	2.5	2	0	1	1
데이터베이스 이용료(\$)	-	-			7,549 (6.68)	904 (0.80)	4,226	4,226
검색시간(분)	3	20	11.5	7.7	3	2.2	2.6	1.8
가공시간(분)	2	6	4		1	1	1	
총비용	865	4,498	2,681	2,681	8,241	1,457	4,849	4,849
적합성1	2.3	1.5	1.9	1.7	1	0	0.5	0.2
적합성2	2.1	1	1.5		0	0	0	

〈표 11〉 질문10의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	S1	S2			S1	S2		
탐색도구	미상무성	AltaVista			GPO Monthly	GPO Pub.		
최종검색건수	3	2	2.5	2.5	15	15	15	15
데이터베이스 이용료(\$)	-	-			7,063 (6.25)	5,764 (5.10)	6,413	6,413
검색시간(분)	10	10	10	7	3	2.1	2.5	1.7
가공시간(분)	5	3	4		1	1	1	
총비용	2,595	2,249	2,422	2,422	7,755	6,300	7,027	7,027
적합성1	2.7	4.5	3.6	3	5	4	4.5	3.2
적합성2	1.3	3.5	2.4		2	2	2	

현황을 조사하고자 한다(통계/수치).

인터넷 검색에서 S1은 미상무성(US Dept of Commerce) 홈페이지로 접속하였고, S2는 AltaVista를 사용하여 US GPO 홈페이지에 접속하였는데 완전한 데이터를 얻을 수 없었다. 한편 DIALOG 검색에서는 S1은 GPO Monthly, S2는 GPO Publication을 사용하였는데 서지 정보만 검색되어 적합도가 낮게 나타났다.

(2) 에너지 관련 분야

① 질문 11: IMF (1997)이후의 한국 천연가스 수급동향을 조사하고자 한다(동향분석정보).

인터넷 검색에서는 S1은 산자부 웹사이트, S2는 OECD 웹 사이트를 이용하였는데 검색건수에 비해 비교적 높은 적합도를 구할 수 있었다. DIALOG 검색에서는 Energyline을 사용하였는데 인터넷에 비해 최종검색 건수는 많았

지만 적합도는 매우 낮게 나타났다.

② 질문 12: 질화 물리브덴 촉매 관련 문헌을 조사하라(書誌정보).

인터넷 검색에서는 검색자 모두가 AltaVista를 사용하였다. DIALOG 검색에서는 CA를 사용하였다.

③ 질문 13: 천연가스 자동차의 연료배출에 관한 학술논문을 조사하고자 한다(全文정보).

인터넷 검색에서는 최종 검색건수가 S1은 AltaVista 사용으로 15건, S2는 Yahoo 사용으로 1건으로 많은 차이를 보여 주었으나 적합성은 모두 '보통'으로 나타났다. DIALOG 검색의 경우 각각 CA, ENERGY SCITEC을 사용하였고 서지 및 전문정보를 검색할 수 있어 매우 높은 적합도를 나타내고 있다.

④ 질문 14: 에너지 연구학자 Gordon Richard L. 의 논문을 조사하라(書誌정보).

인터넷 검색에서는 두 검색사 모두 일부 서지

〈표 12〉 질문11의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	S1	S2			S1	S2		
탐색도구	산자부	OECD			Energyline	Energyline		
최종검색건수	1	1	1	1	23	21	22	22
데이터베이스 이용료(\$)	-	-			39,997 (35.39)	38,698 (34.24)	39,347	39,347
검색시간(분)	4	3	3.5	3	1.2	1.4	1.3	1.6
가공시간(분)	3	2	2.5		2.1	1.9	2	
총비용	1,211	865	1,038	1,038	40,567	39,268	39,917	39,917
적합성1	5	4	4.5	4.2	1.1	1.6	1.3	1.2
적합성2	4	4	4		1	1.4	1.2	

〈표 13〉 질문12의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	AltaVista	AltaVista			CA	CA		
최종검색건수	21	29	25	25	46	54	50	50
데이터베이스 이용료(\$)	-	-			139,568 (123.49)	102,328 (90.54)	120,948	120,948
검색시간(분)	3	6	4.5	7.2	2.2	8.3	5.2	4.7
가공시간(분)	9	11	10		3.9	4.6	4.2	
총비용	1,903	2,941	2,422	2,422	140,623	104,559	122,591	122,591
적합성1	2.7	2.3	2.5	2.3	4.2	4.1	4.1	4
적합성2	2.1	2.3	2.2		3.9	3.9	3.9	

〈표 14〉 질문13의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	AltaVista	Yahoo			CA	ENERGY SCITEC		
최종검색건수	15	1	8	8	48	41	44	44
데이터베이스 이용료(\$)	-	-			145,287 (128.55)	125,079 (110.67)	135,183	135,183
검색시간(분)	15	10	12.5	9	9.8	7.9	8.8	7.7
가공시간(분)	8	3	5.5		7.2	6.1	6.6	
총비용	3,979	2,249	3,114	3,114	148,228	127,501	137,864	137,864
적합성1	3.8	3	3.4	3.2	4.6	4.2	4.4	4.2
적합성2	3.2	3	3.1		4.2	4.1	4.1	

정보만을 확인할 수 있었으나 DIALOG 검색에서는 ENERGY SCITEC, Energyline을 사용하여 비교적 정확한 서지정보를 검색할 수 있었다.

⑤ 질문 15: 촉매연소(Catalytic heater)에 관한 특허를 조사하라(특허정보).

인터넷 검색의 경우 검색자 모두 IBM 특허 웹

사이트를 사용하여 정보를 검색하였다. DIALOG 검색에서는 검색건수가 많아 미국특허 데이터베이스로 제한하여 검색하였다.

⑥ 질문 16: 일본 동경가스(東京ガス) 주식회사의 연차보고서(アニュアルレポート)를 입수하고자 한다(비영어권).

〈표 15〉 질문14의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	Yahoo	AltaVista			ENERGY SCITEC	Energyline		
탐색도구	Yahoo	AltaVista			ENERGY SCITEC	Energyline		
최종검색건수	22	15	18.5	18.5	6	17	11.5	11.5
데이터베이스 이용료(\$)	-	-			10,951 (9.69)	29,690 (26.27)	20,320	20,320
검색시간(분)	6	8	7	6.2	0.9	2.6	1.7	1.4
가공시간(분)	8	3	5.5		1.2	1.1	1.1	
총비용	2,422	1,903	2,162	2,162	11,314	30,330	20,822	20,822
적합성1	2.7	2.2	2.4	2	3.6	4.9	4.2	4.2
적합성2	2.3	1	1.6		3.9	4.7	4.3	

〈표 16〉 질문15의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	IBM	IBM			CLIMS/US PATENT	CLIMS/US PATENT		
탐색도구	IBM	IBM			CLIMS/US PATENT	CLIMS/US PATENT		
최종검색건수	32	32	32	15	65	57	61	61
데이터베이스 이용료(\$)	-	-			137,274 (121.46)	119,552 (105.78)	128,413	128,413
검색시간(분)	5	7	6	6	2.2	2	2.1	2.9
가공시간(분)	6	6	6		4	3.7	3.8	
총비용	1,903	2,249	2,076	2,076	138,346	120,538	129,442	129,442
적합성1	3.9	3.6	3.7	3.6	4.9	4.2	4.5	4.5
적합성2	3.8	3.2	3.5		4.7	4.6	4.6	



〈표 17〉 질문16의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	S1	S2			S1	S2		
탐색도구	동경가스	동경가스			Teikoku Databank	Energyline		
최종검색건수	1	1	1	1	1	1	1	1
데이터베이스 이용료(\$)	-	-			74,401 (65.83)	2,599 (2.3)	38,500	38,500
검색시간(분)	7	4	5.5	4	2.4	0.8	1.6	1.1
가공시간(분)	3	2	2.5		0.9	0.6	0.7	
총비용	1,730	1,038	1,384	1,384	74,971	2,841	38,906	38,906
적합성1	5	4.6	4.8	4.7	4.2	3.8	4	4.1
적합성2	5	4.3	4.6		4.9	3.8	4.3	

〈표 18〉 질문17의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	S1	S2			S1	S2		
탐색도구	AltaVista	OECD			Energyline	ENERGY SCITEC		
최종검색건수	3	1	2	2	28	16	22	22
데이터베이스 이용료(\$)	-	-			48,542 (42.95)	31,408 (27.79)	39,975	39,975
검색시간(분)	10	3	6.5	5	2.7	2.3	2.5	2.2
가공시간(분)	5	2	3.5		2.1	1.8	1.9	
총비용	2,595	865	1,730	1,730	49,372	32,117	40,744	40,744
적합성1	3.1	3	3.0	3.0	2.2	3	2.6	2.3
적합성2	3.3	3	3.1		2.1	2.1	2.1	

인터넷 검색에서 두 명의 검색사 모두가 동경 가스(東京ガス) 웹사이트에 접속하여 검색하였다. DIALOG 검색에서 S1의 경우 영문으로 된

동경가스 연차보고서에 준 하는 데이터를 검색할 수 있었지만, 인터넷에서와 같이 만족할만 한 정보는 검색하지 못했다.

〈표 19〉 질문18의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	IIGU	IIGU			Eventline	Energyline		
최종검색건수	1	1	1	1	7	12	9.5	9.5
데이터베이스 이용료(\$)	-	-			11,911 (10.54)	21,665 (19.17)	16,788	16,788
검색시간(분)	15	5	10	6	2.2	2.9	2.5	2.0
가공시간(분)	2	2	2		1.1	2.1	1.6	
총비용	2,941	1,211	2,076	2,076	12,481	22,530	17,505	17,505
적합성1	4	4.6	4.3	4.2	1.3	1.7	1.5	1.4
적합성2	4	4.2	4.1		1.6	1.2	1.4	

〈표 20〉 질문19의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	AltaVista	AltaVista			ENERGY SCITEC	ENERTY SCITEC		
최종검색건수	52	1	25.5	25.5	20	19	19.5	19.5
데이터베이스 이용료(\$)	-	-			36,776 (32.54)	35,330 (32.26)	36,053	36,053
검색시간(분)	5	6	5.5	5.2	11.1	10.5	10.8	6.4
가공시간(분)	8	2	5		2	2	2	
총비용	2,249	1,038	1,643	1,643	39,042	37,492	38,267	38,267
적합성1	2.1	2	2.0	1.9	4	3.9	3.9	3.8
적합성2	1.8	2	1.9		3.8	3.8	3.8	

⑦ 질문 17: OECD국가의 천연가스 생산현황을 10년간 소급하여 조사하고자 한다(수치/통계). 인터넷 검색에서 검색자 S2의 경우에는

OECD 홈페이지에 접속하여 정보를 검색하였지만 10년간 소급데이터는 구할 수 없었다. DIALOG 검색에서도 소급 데이터를 구할 수 없

〈표 21〉 질문20의 검색결과

구 분	인터넷		평균값		DIALOG		평균값	
	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
검색자	S1	S2	평균	종합 평균	S1	S2	평균	종합 평균
탐색도구	AltaVista	AltaVista			Nuclear Sci. Abs.	ENERGY SCITEC		
최종검색건수	15	12	13.5	13.5	3	48	25.5	25.5
데이터베이스 이용료(\$)	-	-			6,690 (5.92)	88,144 (77.99)	47,417	47,417
검색시간(분)	4	8	6	6	4.1	3.7	3.9	2.9
가공시간(분)	7	5	6		1.6	2.5	2.0	
총비용	1,903	2,249	2,076	2,076	7,676	89,216	48,446	48,446
적합성1	0	3	1.5	1.2	4.2	3.8	4	3.9
적합성2	0	2	1		3.9	3.7	3.8	

었으며 일부 기사 정보를 검색하였다.

⑧ 질문 18: 2000년 천연가스 국제회의 개최 안내 정보를 입수하고자 한다(회색문헌/회의정보).

인터넷 검색에서 두 검색사 모두 국제가스기구(IIGU) 웹사이트를 사용하였는데 적합도는 비교적 높게 나타났다. DIALOG에서는 두 경우 Eventline과 Energyline을 사용하였는데 최신성 부족으로 만족할 만한 정보는 검색할 수 없었다.

⑨ 질문 19: LNG 저장탱크 파열에 관한 문헌을 조사하라(全文정보).

인터넷 검색에서는 S1의 경우 최종검색 건수가 높았지만 비적합 자료가 많아 적합도는 낮게 나타났다. DIALOG 검색의 경우 검색사 모두 ENERGY SCITEC을 사용하였고 균등한 書誌정보를 검색할 수 있었으나 全文정보를 검색 할 수 없었기 때문에 적합도는 예상만큼 높지 않았다.

⑩ 질문 20: 핵연료의 소결체를 인출하는 장치와 관련된 자료를 조사하라(書誌정보).

인터넷의 경우 학술논문의 書誌정보는 거의 검색할 수가 없었는데 이는 원자력분야의 특수성 때문으로 판단되어진다. DIALOG 검색에서는 Nuclear Sci. Abs.와 ENERGY SCITEC을 사용하였는데 적합도가 높은 자료를 검색할 수 있었다.

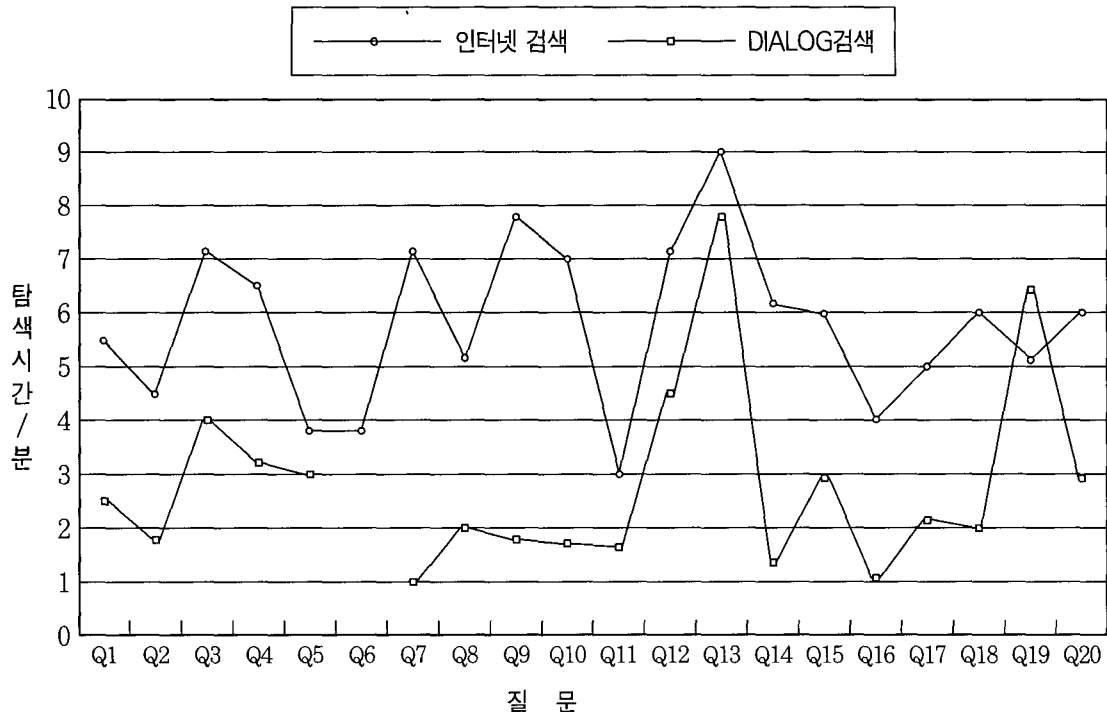
### 3.2.3.2 검색효율 측정

검색효율의 측정은 검색결과와 정확율과 재현율로 측정하는 것이지만 본 연구에서는 인터넷과 DIALOG에서 재현율의 측정이 적절치 않다고 판단되어 두 검색시스템의 탐색 소요시간, 비용, 적합도를 기준으로 측정하고자 한다

#### 1) 탐색 소요시간

DIALOG는 데이터베이스가 정형화되어 있고 중복자료를 자동으로 제거할 수 있기 때문에 탐색 소요시간이 빠르다는 장점을 가지고 있다. 〈표 22〉에서 인터넷과 DIALOG의 탐색 소요시간을 비교해 보았다. 20개의 질문을 검색하는데 소요된

〈표 22〉 탐색 소요시간 비교



시간은 인터넷 검색은 평균 5.8분, DIALOG 검색은 평균 2.8분으로 인터넷 검색이 DIALOG 검색에 비해 2배 정도 더 걸리는 것으로 나타났다.

주제분야별로 탐색소요시간을 살펴보면 경제학분야의 질문1(Q1)-질문10(Q10) 중 인터넷은 평균 5.82분, DIALOG는 평균 2.34분이 소요되었고, 에너지분야의 질문 11(Q11)-질문 20(Q20)에서는 인터넷은 평균 5.76분, DIALOG는 평균 3.28분이 소요되었다. 따라서 경제학분야는 약 2배, 에너지분야는 약 3배의 차이를 나타내고 있다. 인터넷의 탐색 소요 시간이 DIALOG에 비해 높은 것은 검색 결과 중에 중복 자료가 많고 정보요구자에게 제공하기 위한 결과물을 편집해야 하기 때문에 가공시간이 많이 걸릴 뿐만 아니라 검색 시간도 대체로 길기 때문이다.

2) 적합성

인터넷 데이터베이스 검색과 DIALOG검색의 적합성을 좀 더 체계적으로 비교 분석하기 위해서 앞부분에서 세운 아홉 개의 가설을 중심으로 정리하였다.

(1) 가설 1

① 가설 1-1: '에너지 분야에서 DIALOG를 활용한 검색 결과의 적합도가 인터넷을 활용한 경우의 적합도보다 높을 것이다'는 가설을 검증하기 위해서 에너지 분야의 질문 10개의 결과를 그대로 이용하지 않았다. 그 이유로는 DIALOG에 강한 것으로 알려진 書誌, 全文에 관한 질문이 6건, 인터넷에 강한 것으로 알려진 통계, 수치, 비영어권 등에 관한 질문이 4건으로 6 대 4의 비율

로 이루어졌기 때문이다.

따라서 가설 검증을 위해 DIALOG에 강한 것으로 알려진 書誌, 全文, 특허에 관한 질문에서 편향이 심하지 않은 질문 2개를 추출하고, 인터넷에 강한 것으로 알려진 동향분석, 통계/수치에서 편향이 심하지 않은 질문 2개를 추출하여 총 4개 질문의 평균 적합도를 계산하였다. 평균 적합도는 인터넷, DIALOG가 각각 3.3563, 3.5688로 나타나 DIALOG의 적합도가 조금 높게 나타났다. 평균 적합도 차이가 통계적으로 유의미한 것임을 증명해 보기 위해 대응-표본 T검정(t-tests for paired samples)을 한 결과 통계적으로 무의미한 것으로 나타나 가설은 검증되지 않았다.

② 가설 1-2: '경제 분야에서 DIALOG를 활용한 검색 결과의 적합도가 인터넷을 활용한 경우의 적합도 보다 높을 것이다'는 가설을 검증하기 위해서 경제 분야의 질문 10개의 결과를 그대로 이용하지 않았다. 그 이유로는 경제학 질문이 DIALOG에 강한 것으로 알려진 書誌, 및 全文에 관한 질문이 2건, 인터넷이 강한 것으로 알려진 통계, 수치, 비영어권 등에 관한 질문이 8건으로 2대 8의 비율로 이루어졌기 때문이다.

따라서 질문의 비율을 맞추기 위해 DIALOG에 강한 것으로 알려진 書誌, 全文에 관한 질문을 각각 하나씩 추출하고 인터넷에 강한 것으로 알려진 동향분석, 통계/수치에 관한 질문을 각각 하나씩 추출하여 총 4개 질문의 평균 적합도를 계산하였다. 인터넷과 DIALOG의 평균 적합도가 각각 2.7438, 3.7875로 나타나 DIALOG의 적합도가 1.04 높게 나타났으며 두 평균값의 차이가 통계적으로 유의미한 것임을 증명해 보기 위해 앞의 경우 처럼 대응-표본 T검증을 한 결과 통계적으로 유의미한 것으로 나타나 가설은 검증되었다.

## (2) 가설 2

① 가설 2-1: '학술잡지의 全文 정보는 인터넷 데이터베이스 보다 DIALOG의 적합도가 높을 것이다' 라는 가설을 검증하기 위해서 학술잡지 全文정보에 관한 질문(질문 4, 질문13, 질문19)들을 추출하여 적합도를 비교해 보았다. DIALOG 검색의 평균 적합도는 4.0로 인터넷 검색의 평균 적합도인 2.55 보다 높게 나타나 가설 2-1은 검증되었다.

② 가설 2-2: '학술잡지의 書誌 정보는 인터넷 데이터베이스 보다 DIALOG의 적합도가 높을 것이다' 라는 가설을 검증하기 위해서 학술잡지 書誌정보에 관한 질문(질문3, 질문12, 질문14, 질문20)들을 추출하여 적합도를 비교해 보았다. DIALOG 검색의 평균 적합도는 4.38로 인터넷 검색의 평균 적합도 2.08 보다 높게 나타나 가설 2-2은 검증되었다.

③ 가설 2-3: '특허 정보는 인터넷 데이터베이스 보다 DIALOG의 적합도가 높을 것이다' 라는 가설을 검증하기 위해서 특허에 관한 질문(질문 15)을 추출하여 적합도를 비교해 보았다. DIALOG 검색의 적합도는 4.5로 인터넷 검색의 적합도 3.6 보다 높게 나타나 가설 2-3은 검증되었다.

④ 가설 2-4: '수치 및 통계 정보는 DIALOG 보다 인터넷 데이터베이스의 적합도가 높을 것이다' 라는 가설을 검증하기 위해서 수치/통계에 관한 질문(질문1, 질문5, 질문7, 질문9, 질문10, 질문17)들을 추출하여 적합도를 비교해 보았다. 인터넷 검색의 평균 적합도는 2.9로 DIALOG 검색의 평균 적합도 1.6 보다 높게 나타나 가설 2-4는 검증되었다.

⑤ 가설 2-5: '동향분석자료는 DIALOG 보다 인터넷 데이터베이스의 적합도가 높을 것이다'

라는 가설을 검증하기 위해서 동향분석에 관한 질문(질문8, 질문11)들을 추출하여 적합도를 비교해 보았다. 인터넷 검색은 평균 적합도는 3.55로 DIALOG 검색의 평균 적합도 2.1 보다 높게 나타나 가설 2-5가 검증되었다.

⑥ 가설 2-6: '회색문헌 및 회의정보는 DIALOG 보다 인터넷 데이터베이스의 적합도가 높을 것이다' 라는 가설을 검증하기 위해서 회색문헌/회의정보에 관한 질문(질문6, 질문18)들을 추출하여 적합도를 비교해 보았다. 인터넷 검색은 평균 적합도 4.4로, DIALOG 검색의 적합도 1.4 보다 높게 나타나 가설 2-6이 검증되었다.

⑦ 가설 2-7: '비영어권 정보는 DIALOG 보다 인터넷 데이터베이스의 적합도가 높을 것이다' 라는 가설을 검증하기 위해서 비영어권에 관한 질문(질문2, 질문16)들을 추출하여 적합도를 비교해 보았다. 인터넷 검색은 평균 적합도 4.85로

DIALOG 검색의 평균 적합도 2.05 보다 높게 나타나 가설 2-7이 검증되었다.

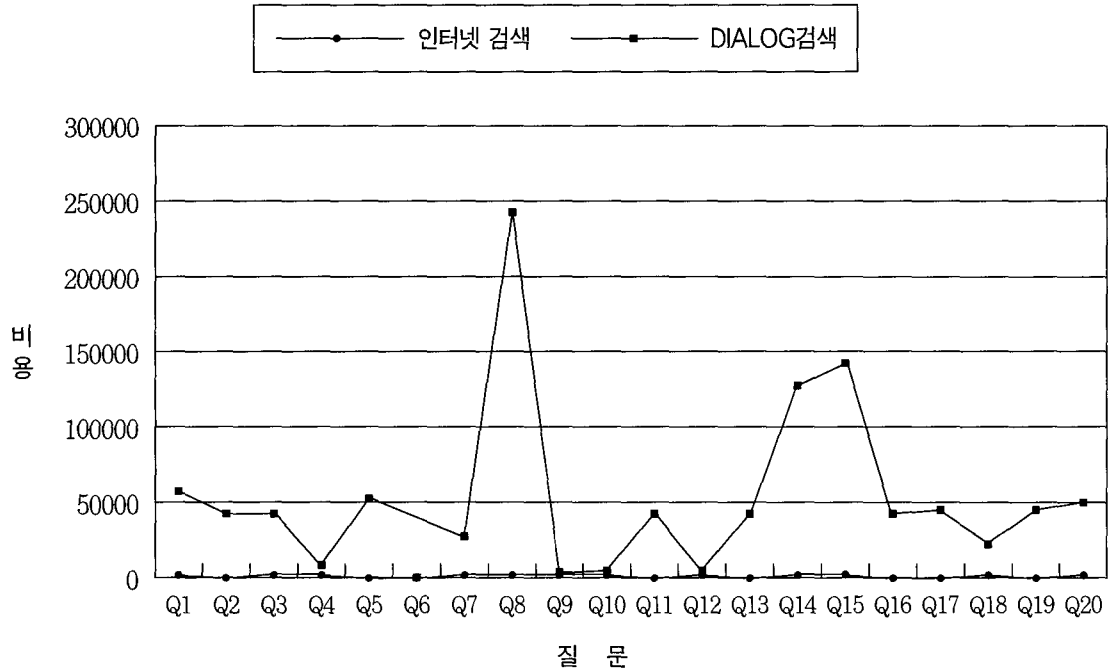
(3) 비용

비용은 인터넷 검색의 경우 탐색소요시간을 기초자료로 하여 검색사 시간 당 비용을 산출하여 측정하였으며 DIALOG 검색의 경우는 인터넷에서 처럼 탐색소요시간을 기초 자료로 하여 검색사 시간 당 비용을 산출하여 측정된 비용에 데이터베이스 이용료를 포함시켰다. 여기에서 데이터베이스이용료는 위탁검색비용 및 부가 의뢰비용은 제외한 이용요금이다.

인터넷의 경우 아직까지 정확한 검색비용 측정에 대한 일정한 기준이 마련되어 있지 않아 본 연구에서는 탐색소요시간을 기준하여 <표 23>에서 인터넷과 DIALOG 검색 비용을 비교하여 보았다.

20개의 질문 모두를 검색하는데 소요된 비용은

<표 23> 인터넷과 DIALOG의 비용 비교



인터넷 검색은 평균 1,959원, DIALOG 검색은 평균 53,354원이 소요되었다. DIALOG 검색은 인터넷 검색 보다 약 27배의 높은 비용이 소요된 것으로 나타났는데 이는 인터넷 검색은 데이터베이스를 무료로 이용할 수 있는 반면 DIALOG는 데이터베이스 이용요금이 고가이기 때문이다. 질문8의 DIALOG 검색에서 EIU(Economist Intelligence Unit)에서 발행하는 고가격의 전문 데이터베이스를 검색하여 239,310원이 소요되었다.

주제분야별로 살펴보면 경제학분야의 질문(Q1-Q11) 중 인터넷 검색은 평균 2,023원, DIALOG 검색은 평균 25,353원이 소요되었고 에너지분야의 질문(Q11-Q20)에서는 인터넷은 평균 2,136원, DIALOG는 평균 53,885원이 소요되었음을 알 수 있다. 인터넷의 소요비용은 주제분야와 관계없이 유사한 값을 나타내고 있는 반면 DIALOG 경우는 에너지 분야가 경제 분야보다 약 2배 정도 많은 비용이 소요되었다.

#### 4 결 론

본 연구에서는 정보시스템과 정보의 유형이 검색 효율에 어떤 영향이 미치는 지를 알아 보기 위해서 아홉 가지 가설을 세워 실험을 통해 검증해 보았다. 연구 결과를 정리해 보면 다음과 같다.

① 에너지와 경제 분야 모두 DIALOG를 활용한 검색 결과의 적합도가 인터넷을 활용한 경우의 적합도 보다 높은 것으로 나타났고 에너지 분야에서는 적합도의 차이가 통계적으로 무의미한 것으로 나타났고 경제 분야에서는 적합도의 차이가 통계적으로 유의미한 것으로 나타났다.

② 학술잡지의 全文 정보는 인터넷 데이터베이스보다 DIALOG의 적합도가 높은 것으로 나타났다.

③ 학술잡지의 書誌 정보는 인터넷 데이터베이스보다 DIALOG의 적합도가 높은 것으로 나타났다.

④ 특허 정보는 인터넷 데이터베이스보다 DIALOG의 적합도가 높은 것으로 나타났다.

⑤ 수치 및 통계 정보는 DIALOG 보다 인터넷 데이터베이스의 적합도가 높은 것으로 나타났다.

⑥ 동향분석자료는 DIALOG 보다 인터넷 데이터베이스의 적합도가 높은 것으로 나타났다.

⑦ 회색문헌 및 회의정보는 DIALOG 보다 인터넷 데이터베이스의 적합도가 높은 것으로 나타났다.

⑧ 비영어권 정보는 DIALOG 보다 인터넷 데이터베이스의 적합도가 높은 것으로 나타났다.

⑨ 탐색 소요시간은 인터넷 검색은 평균 5.8분, DIALOG 검색은 평균 2.8분으로 인터넷 검색이 DIALOG 검색에 비해 2배 정도 더 걸리는 것으로 나타났다.

⑩ 검색하는데 드는 총비용은 인터넷 검색은 평균 1,959원, DIALOG 검색은 평균 53,354원이 소요되어 DIALOG 검색이 인터넷 검색 보다 약 27배의 높은 비용이 소요되었다.

종합해 보면 정보 유형이 DIALOG와 인터넷 데이터베이스 시스템의 적합성에 영향을 미칠 것이라는 가설들은 별 무리없이 검증되었으나 DIALOG를 활용한 검색 결과의 적합도가 인터넷을 활용한 경우의 적합도 보다 높을 것이라는 가설은 에너지 분야는 통계적으로 검증되지 못했고 경제 분야에서만 검증되었다. 가설을 설정할 당시에는 아직까지는 거의 모든 주제 분야에서 DIALOG의 검색 효율이 인터넷에 비해 훨씬 높을 것이라고 가정하였는데 실험 결과에서는 두 분야에서 DIALOG가 약간 적합도가 높은 것으로 나타난 것은 의외의 결과였다. 이러한 결과가 나온 이유들을 정리해 보면 첫째는 인터넷 검색에

서 專門 정보가 비교적 많이 수록된 인터넷 專門 검색엔진을 사용하게 한 점을 들 수 있고, 둘째는 인터넷 데이터베이스의 질적인 성장을 들 수 있겠고, 마지막으로 경제 및 에너지 분야의 주제의

특성상 인터넷 데이터베이스를 통해 관련 자료를 많이 얻을 수 있지 않았을까 하는 가정을 해 볼 수 있다.

## 참 고 문 헌

- 橋田昌明. 1998. 商用DB 檢索システムと比較したの檢索エンジン. 『專門圖書館』, No.171: 61-67.
- 김현희, 안태경. 1999. 인터넷 데이터베이스와 온라인 데이터베이스의 검색시스템 검색효율 비교. 『文獻情報學論集(明知大)』, 第6號: 275-301.
- Anagnostelis, Betsy. 1998. Evaluation criteria for different versions of the same database: a comparison of Medline services available via the World Wide Web. *Online information 97. 21st International Online Information Meeting Proceedings, 9-11 December 1997*. [Lodon: Learned Information Europe Ltd.].
- Arnold, Stephen E. and Arnold Erik S. 1998. Search engines: lost and found in cyberspace. *Proceedings of the 19th National Online Meeting, 12-14 May 1998*. [New York : Information Today Inc.].
- Bates, Mary Ellen. 1999. *Super searchers do business: the online secrets of top business researchers*. New Jersey: CyberAge Books.
- Chu, Heting. 1998. Internet search services vs. online database service. *Proceedings of the 19th National Online Meeting, 12-14 May 1998*. [New York: Information Today Inc.].
- Feldman, Susan. 1998. The internet search-off: results and ruminations. *Proceedings of the 19th National Online Meeting, May 12-14, 1998*. [New York: Information Today Inc.].
- . 1998. The internet search-off. *Searcher* 6(6) 1998: 28-35
- Harter, Stephen P. and Hert Carol A. 1997. Evaluation of information retrieval systems: approaches, issues, and methods. *Annual Review of Information Science and Technology*, 32: 3-98.
- Hock, Randolph. 1999. *The extreme searcher's guide to Web search engines: a handbook for the serious searcher*. New Jersey: CyberAge Books.
- Jacso, Peter. Mapping algorithms to translate natural language questions into search queries for web database. *Proceedings of the 18th National Online Meeting, 13-15 May 1997*. [New York:



- Information Today Inc].
- Jacobs, Paul S. 1998. Browsing vs. surfing: the next generation of search and retrieval. *Proceedings of the 19th National Online Meeting, 12-14 May 1998*. [New York: Information Today Inc.].
- LaChance, Sue. 1998. Consumer-oriented information retrieval: informed communities find buying power on the net. *Proceedings of the 19th National Online Meeting, 12-14 May 1998*. [New York: Information Today Inc.].
- Liddy, Elizabeth D. 1998. Beyond retrieval. *Proceedings of the 19th National Online Meeting, 12-14 May 1998*. [New York: Information Today Inc.].
- Platt, Nina. 1998. Search engines for intranets. *Proceedings of the 19th National Online Meeting, 12-14, May 1998*. [New York: Information Today Inc.]
- Prestamo, Anne Tubbs. Development of web-based tutorials for online database. *Proceedings of the 19th National Online Meeting, 12-14 May 1998*. [New York: Information Today Inc.].
- Rieb, Soo Young. 1998. User-computer-librarian interaction in end-user online searching. *Proceedings of the 19th National Online Meeting, 12-14 May 1998*. [New York: Information Today Inc.].
- Spink, Amanda and Bateman Judy. 1998. User's searching behavior on the exte web search engine. *Proceedings of the 19th National Online Meeting, 2-14 May 1998*. [New York: Information Today Inc.].
- Su, Louise T. 1997. Developing a comprehensive and systematic model of user evaluation of web-based search engines. *Proceedings of the 18th National Online Meeting, 13-15 May 1997*. [New York: Information Today Inc.].
- Tragert, Joseph. 1997. CD-ROM vs. online vs. internet: continuous change. *Proceedings of the 18th National Online Meeting, 13-15 May 1997*. [New York: Information Today Inc.].
- Weiner, Michael L. and Rusch, Peter F. 1997. New searching technologies and interfaces. *Proceedings of the 18th National Online Meeting, 13-15 May 1997*. [New York: Information Today Inc.].
- Williams, Martha E. 1998. Highlights of the online database industry and the internet. *Proceedings of the 19th National Online Meeting, 12-14 May 1998*. [New York: Information Today Inc.].
- Xie, Hong(Iris). 1998. Access to online database: an exploration of user's experiences with web interfaces. *Proceedings of the 19th National Online Meeting, 12-14 May 1998*. [New York Information Today Inc.].